A quick overview of the Swendsen-Wang method and its application to Boltzmann machines

Simon Osindero

University of Toronto

Tea Talk: 14-July-2004

Introduction

- Several efficient Monte Carlo methods involve augmenting the original variables in a model with a set of auxiliary variables.
- Operating in this enlarged space can sometimes be much easier and more efficient than operating in the original space.
  - One example of this type of method is the Hybrid Monte Carlo approach for systems of continuous variables.
  - Swendsen-Wang methods (and their generalisations) are another example, and have been mostly applied to systems of discrete variables.

Motivation

- Sampling from Boltzmann machines (and similar MRF models) is difficult in the general case. Markov Chain Monte Carlo (MCMC) methods are often required.
- Unless we have particular structure (eg: a bipartite graph as in Restricted Boltzmann Machines), the method commonly used is single variable updating using Gibbs sampling.
  - This may suffer from slow mixing, requiring many updates to produce significant changes in the global state.
- Enhanced methods that allow multiple variables to be updated simultaneously can potentially lead to faster mixing and greater efficiency.

Basic Idea: Introduce Auxiliary Variables

- Somewhat surprisingly, we can sometimes make a sampling problem easier by adding more variables.
  - Interested in samples from $P(\mathbf{x})$.
  - Introduce auxiliary variables u_k.
  - Form the joint distribution $P(\mathbf{x}, \mathbf{u}) = P(\mathbf{x})P(\mathbf{u}|\mathbf{x})$.
  - Carefully select/design $P(\mathbf{u}|\mathbf{x})$ so that:
    - ★ $P(\mathbf{u}|\mathbf{x})$ is easy to sample from;
    - ★ $P(\mathbf{x}|\mathbf{u})$ is also easy to sample from.
- Sample from the joint model by alternately sampling these conditionals.
- The marginals for $\mathbf{x}$ match our desired distribution.

Generalised Swendsen-Wang

- Desired distribution: $P(\mathbf{x}) = \frac{1}{\mathcal{Z}_x}\pi_0(\mathbf{x}) \prod_k \Phi_k(\mathbf{x})$.
  - $\pi_0(\mathbf{x})$ is a simple base density, perhaps factorised over $\{x_i\}$.
- Introduce k auxiliary variables u_k, one for each Φ_k.
- Choose the following form for their conditional density,

$$P(\mathbf{u}|\mathbf{x}) = \prod_k \frac{1}{\Phi_k(\mathbf{x})}\mathcal{I}\left[0 \leq u_k \leq \Phi_k(\mathbf{x})\right] \tag{1}$$

  where $\mathcal{I}$ is an indicator function which is 1 when its argument is true, and zero otherwise.
- Each u_k is independently and uniformly distributed from 0 to $\Phi_k(\mathbf{x})$.

Generalised Swendsen-Wang (cont'd)

- Our choice of $P(\mathbf{u}|\mathbf{x})$ leads to the following joint distribution,

$$P(\mathbf{x},\mathbf{u}) = \frac{1}{\mathcal{Z}_x}\pi_0(\mathbf{x})\prod_k \mathcal{I}\left[0 \le u_k \le \Phi_k(\mathbf{x})\right] \qquad (2)$$

  - Note that the Φ_k terms from $P(\mathbf{x})$ and from $P(\mathbf{u}|\mathbf{x})$ have canceled.

- We also have

$$P(\mathbf{x}|\mathbf{u}) \propto \pi_0(\mathbf{x})\prod_k \mathcal{I}\left[0 \le u_k \le \Phi_k(\mathbf{x})\right] \qquad (3)$$

I.e. $P(\mathbf{x}|\mathbf{u})$ is just the 'base' density $\pi_0(\mathbf{x})$, restricted to the region satisfying the constraints $\{0 \le u_k \le \Phi_k(\mathbf{x})\}$.

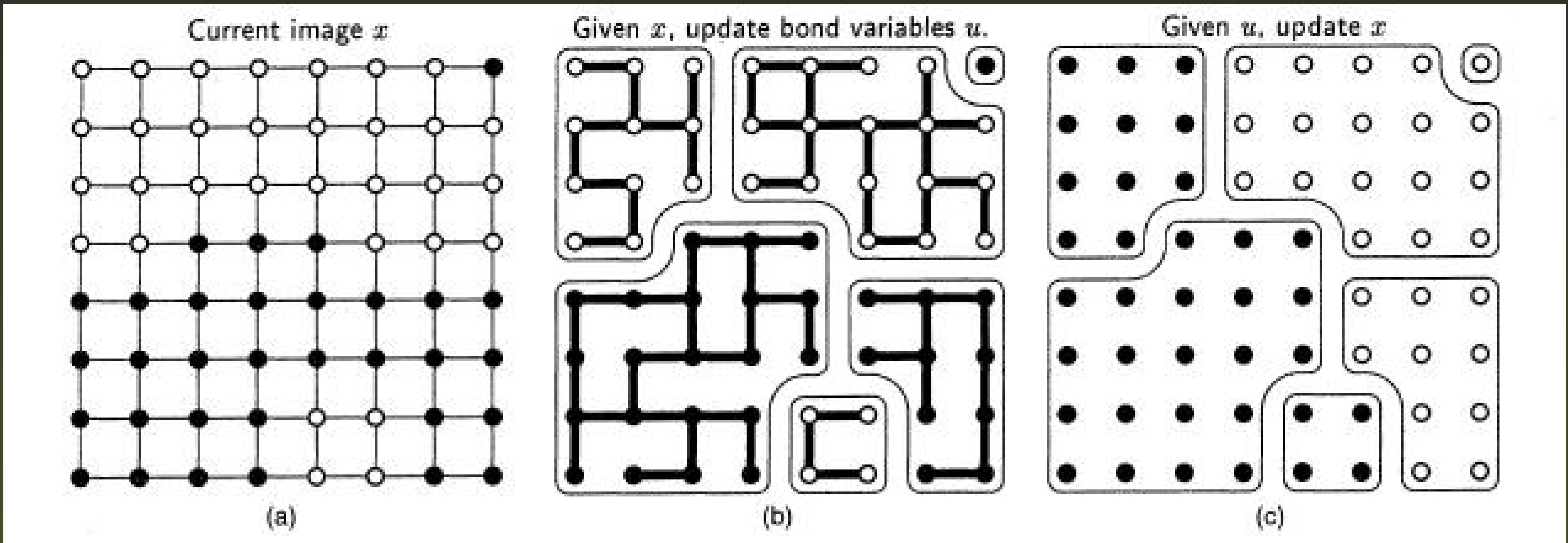
SW Method for [-1,1] Boltzmann Machines

- Potentials $\Phi_k(x_i, x_j)$ can only take two values. $\Phi_k(x_i, x_j) = e^{W_{ij}}$ if $x_i = x_j$, and $\Phi_k(x_i, x_j) = e^{-W_{ij}}$ if $x_i = -x_j$.
- Terms $\mathcal{I}\left[0 \leq u_k \leq \Phi_k(\mathbf{x})\right]$ may constrain the allowed combinations of $\{x_i\}$ when conditioning on $\mathbf{u}$.
- $W_{ij} > 0$
  - If $u_k > e^{-W_{ij}}$ then we must have $x_i = x_j$.
  - If $u_k \leq e^{-W_{ij}}$ then there is no direct constraint on (x_i, x_j).
- $W_{ij} < 0$
  - If $u_k > e^{W_{ij}}$ then we must have $x_i = -x_j$.
  - If $u_k \leq e^{W_{ij}}$ then there is no direct constraint on (x_i, x_j).
- Constraints give rise to “bound” clusters which act as a single unit.
- We can replace $\{u_k\}$ with binary summary variables $b_k = \mathcal{I}[u_k > e^{-|W_{ij}|}]$ — denoting presence of of a “bond”.

"Bond" Variables

- We have one binary bond variable, b_k, for each weight W_{ij} in the original Boltzmann machine.
- Update bonds given $\mathbf{x}$:
  - If $W_{ij} > 0$, and if $x_i = x_j$ then form bond ($b_k = 1$) with probability $p = 1 - e^{-2W_{ij}}$, otherwise $b_k = 0$.
  - If $W_{ij} > 0$, then 'do the opposite'.
- If $b_k = 1$, then the units at the bond terminals must be:
  - in the *same* states if $W_{ij} > 0$;
  - in *different* states if $W_{ij} < 0$.
- Update $\mathbf{x}$ given bonds:
  - Identify connected cluster (units linked by active bonds).
  - Each of these clusters has just two possible states. Select a state based on the probilities that arise from the contribution of the residual $\pi_0(\mathbf{x})$ term — in a BM model this is the combined influence of the single unit biases.

Simple Example Of Update Procedure



- Each unit is connected to 4 nn with weight $W > 0$.
- Bonds form between units in the same state with probability $1 - e^{-2W_{ij}}$.
- After sampling the bonds, we have 6 connected clusters. Each cluster of units is independently resampled as a single block.
- New state differs considerably from the previous state.

Successive Realisations From An Ising Model

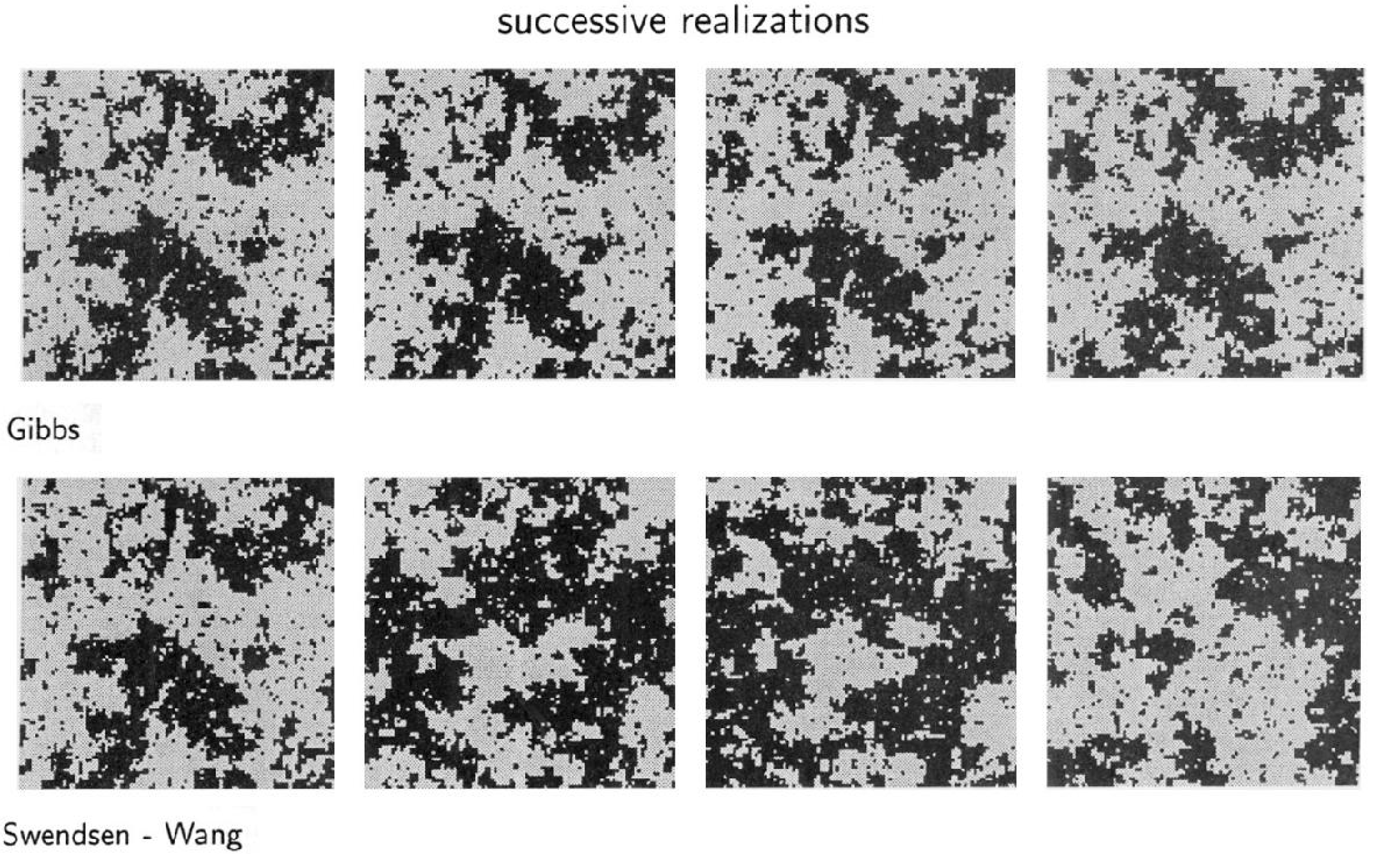


Figure 3. Successive Realizations of the Ising Model at Critical Temperature From the Single-Site Metropolis and Swendsen–Wang Algorithms.

References

- Swendsen, R.H. & Wang, J. *Nonuniversal Critical Dynamics in Monte Carlo Simulations.* Phys. Rev. Lett. **38**(2), 1987.
- Edwards, R.G. & Sokal, A.D. *Generalization of the Fortuin-Kastelyn-Swendsen-Wang representation and Monte Carlo algorithm.* Phys. Rev. D **38**(6), 1988.
- Higdon, D.M. *Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications* Journal of the American Statistical Association **93**(442), 1998. (Figures were cut from this paper.)

- There are also extended methods, for example "Partially Decoupled" SW (Higdon,93&98)